Winter 12-5-2005

# Effective Knowledge Representation Through Data Modelling Approaches

Mik Kim

Wusheng Zhang

# Effective Knowledge Representation Through Data Modelling Approaches

Mik Kim[1], Wusheng Zhang [2]
Centre for Hospitality and Tourism Research[1]
Centre for International Corporate Governance Research[2]
Victoria University, Melbourne, Australia
Email: mik.kim@vu.edu.au wusheng.zhang@vu.edu.au

**Abstract:** Data modelling can be seen as knowledge representation in terms of sharing the same philosophical assumptions. In data modelling process, the recognition of the philosophical background on human inquiry and the nature of knowledge pertinent for appreciating the problems is important as different ontological views lead to different conceptions about data models. Recognising and incorporating different forms of organisational knowledge are also important in the data modelling process as a formal representation of some subset of the knowledge, which the organisation needs to carry out its business. This paper discusses the two distinct philosophical foundations for the effective representation of organisational knowledge.

## I. Introduction

Information systems with information technology have played an important role in capturing, storing, sharing and disseminating organisational knowledge. In the information systems development stage data modelling is an indispensable part of managing organisational knowledge in describing, organising and analysing data that are stored and manipulated in the information systems. Data modelling is an approach that is used to manage complex problems [6]. Data modelling approaches are divided into two paradigms. They are objectivism and subjectivism. Objectivism in data modelling follows the empirical analytical scientific method. It assumes that reality can be described by independent facts, which corresponds to the empirical base of observational statement in objectivist philosophy. A database captures these facts and data models provide the structure for organising all the facts into a consistent picture of reality. In accordance with this, objectivist data modelling approaches were called fact-based.

In contrast, subjectivism of data modelling pursues subjectivist paradigm. It assumes that that the domain of inquiry is not independent of the observer, and therefore reality cannot be described in terms of independent fact. Rather, what counts as reality are socially constructed images, which emerge in social interaction, in particular through communication in some language. The details of these images are not completely arbitrary, but depend on the grammar, the rules and meanings, which govern social communication. Accordingly, a data model attempts to formalise some of the informal social rules and meanings. The subjectivist data modelling approaches were called rule-based.

As Klein and Hirschheim claim that "no data modelling approach can avoid philosophical assumptions because data modelling is a process of inquiry that has intrinsic similarities with classic scientific theory construction" [12, p. 8], understanding philosophical basis is essential to data modellers because different ontological views lead to different conceptions about data models. It is therefore significant for data modellers to be aware of different philosophical assumptions of conceptual data modelling. The purpose of this paper is to explore the implication of two different objectivist and subjectivist's paradigms to conceptual data modelling process to effectively represent organisational knowledge. The organisation of the paper is as follows. First it discusses a philosophical perspective on data modelling in general. After that it moves to two sets of assumptions of data modelling in terms of ontology and epistemology. Following the discussion of the two assumptions, it examines objectivist-subjectivist data modelling. Then, it speculates the fact-based and rule-based of data modelling. Finally, it ends with conclusions and future work.

## II. Philosophical Perspectives on Data Modelling

There is a slight difference between data modelling and conceptual data modelling. Conceptual data modelling is the phase of the development process that involves the abstraction and representation of the real world data pertinent to an organisation [2] whereas data modelling is the activity of creating a data model. Olle [14, p. 46) points out, " A data modelling facility has to be seen as functionally more restricted than a conceptual modelling facility". He also points out that conceptual modelling is much less widely used and a data modelling facility is restricted to data and the processes. Although there are discrepancies, in this paper the term data modelling and conceptual data modelling is used interchangeably.

There are linkages between the concerns of data modelling and philosophy. One result of the philosophical debate on the meaning of language and hermeneutics is that it is impossible to capture knowledge simply in terms of articulable facts as the positivists had claimed. A broader concept of knowledge is relevant for understanding practical difficulties with determining the system or formalisation boundary in specifying computer applications. It is widely

Proceedings of the Fifth International Conference on Electronic Business, Hong Kong, December 5-9, 2005, pp. 582- 586.

recognised that data modelling should include the meanings and relationships among data that may be hidden from the users. It is for this reason that concern for the adequate recognition of the importance of tacit knowledge arises. Tacit knowledge is personal knowledge that is hard to formalise or articulate [15]. Hidden meanings and knowledge can be imperfectly represented during data modelling as it falls within the domain of tacit knowledge.

Moreover, just because users may know how to do their work, it does not follow that it is possible to capture their knowledge in explicit rules and descriptions [16]. This remains true even if some of the users are asked to participate the information systems design teams. In this sense it may be said that data modelling involves the design of a knowledge representation schema, which needs good techniques to decide what to include and what not to include, understanding the important role of tacit knowledge. In addition to the problem of hidden and tacit knowledge, data modelling constructs a formal representation of some subset of the knowledge, which the organisation needs to carry out its business.

Data modelling and knowledge representation can be seen essentially similar in terms of sharing the same fundamental assumptions. The existing approaches of data modelling with regard to the assumptions are about the nature of reality and the nature of the knowledge captured in a data model including the ways and means of how it is collected. That is the paradigmatic dimension of objectivism-subjectivism. The objectivist-subjectivist dimension of a paradigm raises two issues such as one dealing with ontology and the other with epistemology of data modelling.

## III. Ontology *V* Epistemology

Ontology is concerned with the essence of things and the nature of the world. It subsumes two extreme positions of realism and nominalism. Realism postulates that the universe comprises objectively given immutable objects and structures [12]. These exist as empirical entities, on their own, independent of the observer's appreciation of them. On the other hand, nominalism is where reality is not a given immutable out there, but is socially constructed [12]. It is the product of the human mind.

The ontological question what is being modelled points to the fundamental assumption of data modelling approaches about the environment or modelling domain. That is about the nature of the Universe of Discourse (UoD). There are many forms of terminology for talking about the ontology of data modelling. Examples of terms used to describe ontology in data modelling are entities, relationships, facts and speech acts. The ontology of data modelling also includes some fundamental assumptions about the nature of the application domain. That is whether there is a single or several user systems or no system at all, whether the primary constituents of each user system are operations, roles, decisions, social action or speech acts.

In contrast to ontology, epistemology is the grounds of knowledge. The term relates to the way in which the world may be legitimately investigated and what may be considered as knowledge and progress. It includes elements concerned with sources of knowledge, structure of knowledge and the limits of what can be known. Two extreme positions are positivism and interpretivism. According to Klein and Hirschheim [12], positivism implies the existence of causal relationships which can be investigated using scientific method whereas interpretivism implies that there is no single truth that can be proven by such investigation. Different views and interpretations are potentially legitimate and the way to progress is not to try and discover the one correct view but to accept the differences and seek to gain insight by a deep understanding of such complexity.

The epistemological question why the result is valid points to the fundamental assumptions that are made on how one can obtain valid interpretations and knowledge about the UoD. Experimental modes of schema construction like prototyping can be compared with the blue-print specification approach as is associated with many versions of the so called waterfall model of schema development [9][3]. The question of how to cope with different types of uncertainty such as problem during data modelling is not the only consideration that is important in comparing experimental or specification approaches, but it is a good example to demonstrate that different methods imply different assumptions about knowledge and inquiry.

In the context of data modelling, ontological assumptions are concerned with the nature of the UoD to be modelled. Epistemological assumptions are concerned with the appropriate approach for inquiry regarding what one needs to know to create the data model and with the cognitive status of the result is a data model true or it is merely an imaginary tale. That is a simplifying design assumption presumed to be valid only for a particular system at a particular time. The existing approaches of data modelling with regard to the assumptions are about the paradigmatic dimension of objectivism versus subjectivism.

## IV. Objectivism *V* Subjectivism

Objectivism postulates that the UoD is comprised of immutable objects and structures that exist as empirical entities. In principle, a model of the UoD ought to exist which is correct independently of the observers' appreciation of it. A data model is true if it is accurately depicts the underlying reality of the UoD. In contrast, subjectivism in data modelling holds that the UoD is a subjective construction of the mind. A data model can best reflect peoples' conventions or perceptions that are subject to negotiated change. Important mechanisms by which subjective experiences take on an objective quality in the minds of individuals are the rules surrounding institutions, tradition as transmitted through artefacts and changing use of language and sedimentation [1]. Sedimentation refers to

the ordering of experiences, which is transmitted subconsciously by virtue of growing up in one segment of society as opposed to another or in different societies altogether [12]. From this view a data model is correct if it is consistent with the perception of the UoD as constructed by institutional programming, sedimentation and tradition. On the contrary, the objectivist view holds that language neutrally depicts reality, which is the same for all regardless of culture and individual perception.

Burrell and Morgan [4] describe the categorisation of assumptions on objectivist/subjectivist and order/conflict dimensions. According to them, the objectivist position is characterised by the realist ontology and positivist epistemology. The objectivist assumes that objects and structures exist as empirical entities independent of human observers and that the appropriate way of acquiring knowledge of the world is by observation and the identification of causal relationships. The subjectivist position holds that scientific method is not appropriate for explaining the social world as different people interpret the world in different ways and any agreement is inter-subjective. The appropriate way to investigate the social world is to recognise multiple realities and to adopt an interpretivist stance. The order view of the world sees order, stability and integration while the conflict view sees coercion, disintegration, and places an emphasis on change.

The two classifications organise four particular positions such as functionalism (objectivism/order), radical structuralism (objectivism/conflict), social relativism (subjectivism/order), and neohumanism (subjectivism/ conflict). The metaphors of doctor, warrior, liberal teacher, and emancipator respectively have symbolised these groupings. The doctor, technical expert sees data modelling as an objective use where entities and relationships can be said to exist independently of people. The data model is a value-free reflection of a singular reality and scientific method is the appropriate way of building a data model. The warrior also perceives data modelling as an objective use, but assumes that there are conflicting interests and is therefore concerned with whose requirements are being addressed. The difference is that the warrior considers data model as a reflection of real world structures but because of socio-economic factors the data model is not value-free which requires the data modeller's support in the struggle against the owners.

The liberal teacher, facilitator considers reality to be socially constructed and that there are as many realities as there are people involved in the development exercise. A good data model is one, which creates a shared understanding so the appropriate way to go about creating such a model is through participation and facilitation. The emancipator would also consider that reality is socially constructed, but recognises the presence of conflict and hence the exercise of power. The emancipator adopts a critical stance and seeks to initiate change that will improve the situation, recognising that there is unlikely to be a consensus and that any data model will be mediated through the exercise of power. A good data model is one, which helps create a social reality that eases cultural domination via a rational discourse that is free from restrictions.

Under a more radical interpretation of the subjectivist approach, data modelling does not merely reflect social consensus perceptions, but it affects the very process of reality construction. This is so because socially transmitted concepts and names direct how reality is perceived and structured. The construction of reality varies with different languages and cultures. As data modelling typically introduces new concepts and ideas in the users world, it intervenes with the very definition of what counts as reality. It changes what is subjectively experienced as an objective reality and its appreciation what may at one point have been accidental or not even noticed may become very important and recorded in detail. In this sense, data modelling is a form of institutional representation and once a data model has been in use for some time it may become part of sedimentation and tradition. Accordingly, data modelling is never neutral, but rather a biased supporter of reality.

The framework of the categorisation suggested by Burrell and Morgan [4] illustrates that is but one way of approaching the world in assuming that a good data model is one that mirrors reality we are working within a positivist tradition. There are more frameworks proposed that show what the measure of a good data model is. Berger and Luckman [1] suggest a framework where subjective meanings become objective objectivation that then act back as they socialise present and future generations. Reality construction is the dialectical relationship of these three moments. Giddens's [8] proposal has advanced much complementarily from the paradigmatic conclusion of the Burrell and Morgan [4] by conceiving the objective and subjective to be present simultaneously, as represented in structuration theory. An important implication of structuration theory is that the objective/subjective distinction should be seen as duality rather than dualism, and that it is not possible to separate out objective components. This may lead to some inconsistencies in their conceptual base. Such inconsistencies are often not reflected, because technical work in the area of data modelling is often unaware of the deeper stipulations and assumptions on which newly proposed methods and tools rest.

Those assumptions are the principles on fact based, rule based, and frame based approaches. Objectivism in data modelling follows the empirical analytical scientific method. It assumes that reality can be described by independent facts, which corresponds to the empirical base of observational statement in objectivist philosophy. A database captures these facts and data models provide the structure for organising all the facts into a consistent picture of reality. In line with this, Objectivist data modelling approaches were called fact-based. Subjectivism of data modelling, in contrast, pursues subjectivist paradigm. It assumes that that the domain of inquiry is not independent of the observer, and therefore reality cannot be described in terms of independent fact. Rather, what counts as reality are socially constructed

images, which emerge in social interaction, in particular through communication in some language. The details of these images are not completely arbitrary, but depend on the grammar, the rules and meanings, which govern social communication. Hence, a data model attempts to formalise some of the informal social rules and meanings. The subjectivist data modelling approaches were called rule-based.

## V. Fact-Based V Rule-Based Data Modelling

The fact-based approach is grounded on concepts that are related to given facts in the sense of existing state of affairs. The fact-based approaches to data modelling tend to follow the objectivist tradition. A fact is typically defined by associating a specific attribute or relationship to a given entity. Under this interpretation, a data model is like a mirror or picture of reality. Reality is a given and out there. Typical examples for such concepts are entity, property/attributes and relationship. Entities have properties or attributes. Both entities and their properties have an objective existence. Entity-based approaches implement von Wittgenstein's [17] picture theory of meaning. That is data correspond to facts and it is these that entity-based approaches seek to model. This is apparent when Kent [11] comments that by focusing on the facts, we obtain a methodology for data analysis and design which is at once simpler and more powerful than other methodologies for data modelling.

The fact-based approach represents the mainstream of the data modelling field. It has the longest history, and it has been intensively researched over the last two decades [5]. The approaches suggested within this approach have been widely applied, and there exists a considerable amount of application experience. Therefore, the fact-based approach is better understood than that of the rule-based approach.

In contrast, the rule-based approaches to data modelling are heavily influenced by the subjectivist tradition. Their proponents see the main task of data modelling as formalising the meaning of messages that are to be exchanged among a professional community such as managers and engineers. The expression of meanings must follow social determined rules, which facilitate the comprehension of what is communicated. They argue that meaning is created within the human mind and related to human purpose or intentions. The latter arise out of an understanding of reality, socially constructed. That is it emerges from social interaction conditioned by social conventions or rules. For example, all computer data ultimately have to be interpreted in terms of their natural language meanings. Hence data can at best convey meaning from someone, but they cannot have any objective meaning.

In the rule-based approach a data model describes rules which govern the uses of signs and symbols in organisational behaviours and thereby attach specific meanings to the organisational vocabulary. Therefore these approaches are largely derived from the concept of a rule. In general, a rule is regarded as a general prescription that governs the generation, meaning and use of linguistic expressions both informal and formal as are included in the information system.

Rule-based approaches form a minority position in the data modelling field. It has been pursued by few researchers [8] and its history can be dated back to the late 1970s and early 1980s. Rule-based data modelling methods are much less developed than their fact-based counterparts. Moreover, these approaches have not been widely applied in their full form outside academia and experience of their usefulness is limited. One reason for this is their inner complexity and highly developed theoretical vocabulary which is difficult to adopt for practising information systems professionals. Data modelling in the rule-based is an important medium for organisations to learn about their communication practice helping detect distortions and inconsistencies in communication.

There is a third approach on data modelling which combines both objectivist and subjectivist principles of data modelling. The approach is also called frame-based because ideas can be captured in mental frames. The basic modelling construct of an object builds on the concept of an imaginary actor as originally proposed by Kay [10]. An actor in Kay's sense is any software object that behaves in predefined ways. Actors can be used to model real objects, but they can also be used to modelling imaginary objects that represent people's ideas, which exist only in Alice in Wonderland. An actor or frame view of data is based on the idea that a description of data with its permissible operations should be combined into an actor or knowledge frame. The frame-based approach is best viewed as a variant of objectivism in data modelling.

Similar to the frame based approach for the purpose of lessening the danger of one perspective only data modelling, an unbounded systems thinking approach as the application of multiple perspectives to data modelling has been proposed by Mitroff and Linstone [13]. The core three perspectives include organisation (O), technical (T) and personal (P). The T perspective is concerned with a scientific world-view, logic, rationality, modelling and analysis claiming of objectivity. The O perspective is concerned with social entities, politics and processes; and the P perspective is of the individual and concerned with power, influence, prestige, learning and beliefs. The flaw of the approach is associated with practical difficulties by realising different forms of knowledge, which requires different knowledge interests and thus need to be considered jointly. In other words, people with different background and tradition tend to follow one particular perspective over other perspectives for which there are no simple rules for balancing the requirements of different perspectives.

A central issue of data modelling is in assuming that data modelling must be either objective or subjective such that data modelling process can be based either upon the notion of an objective reality or upon the notion of a reality that is socially constructed. In other words, the problem is not whether to be objective or subjective but that one implies the

other and that any method that is based upon one of these ontological positions must accept and cater for the other. Both objective and subjective aspects are present at the same time. Objective methods can be said to be objective because they are not subjective. However, taking a strong objectivist position is difficult. We cannot demonstrate the existence of an objective reality. A strong subjectivist position is another danger for good data modelling. In addition, the failure of information systems development is caused by overlooking many different forms of organisational knowledge. That is beyond of technical aspects. Accordingly, data modelling process should be done by recognising and incorporating the different forms of knowledge, which may influence the success of the entire data modelling process.

## VI. Conclusions And Further Study

Data modelling as an organisational knowledge representation is concerned with the fundamental assumptions. The existing approaches of data modelling with regard to the assumptions are about the nature of reality and the nature of the knowledge captured in a data model. In data modelling process the recognition of paradigmatic dimension of objectivism-subjectivism is important. However, it is not appropriate to consider the data modelling process as objective the notion of an objective reality, or as subjective the notion of a reality that is socially constructed. Without the notion of objectivity the notion of subjectivity is meaningless. It should be noted that the sharp distinction between fact-based and rule-based data modelling approaches is not necessary. Although the major emphasis in the information systems research community has been on the fact-based approaches, data modelling processes need to incorporate some of both fact-based and rule-based approaches and thereby engage in some middle ground between the extremes of fact-based and rule-based. For the effective representation of organisational knowledge, consideration should be given to casting both objectivism and subjectivism as a duality. In addition to taking the duality into consideration, data modelling process should also be done by recognising and incorporating different forms of organisational knowledge. The typical response to the problem of tacit knowledge is the modellers' purpose in applying a data model as a formal representation of some subset of the knowledge, which the organisation needs to carry out its business. Misunderstanding is connected to the failure to understand the important role of tacit knowledge. There is a need for good techniques to decide what to include and what not to include in the subset.

Further study should, therefore, include developing applicable techniques for decision making in selecting organisational tacit knowledge. It should also consider in the area of rule-based approaches of subjectivism, as the concepts of the subjectivist approaches are not well researched in comparison with the fact based objectivism. Moreover, although the fact-based approaches are a lot better researched there is more study needed in terms of development of data modelling approaches that embed the philosophical assumptions.

## References

[1] Berger, P., & Luckmann, T. (1967). *The Social Construction of Reality: A Treatise in the Sociology of Knowledge.* New York: Doubleday.
[2] Brodie, M. L. (1986). On the development of data models. In M. L. Brodie., J. Mylopoulas & J. W. Schmidt (Eds.), *On Conceptual Modelling* (pp. 19-47). New York: Springer-Verlag.
[3] Bubenko, J. A. (1980). Information Modelling in the Context of System Development. In H. A. Lavington. (Ed.), *Proceedings of the IFIP World Congress* (pp.395-411). Amsterdam: North-Holland.
[4] Burrell, G., & Morgan, G. (1979). *Sociological Paradigms and Organisational Analysis.* London: Heinemann Educational Books.
[5] Chen, P. (1981). Entity-Relationship Approach to Systems Analysis. *In Proceedings of the Second International Conference on E-R Approach* (pp. 12-14). Washington: E-R Institute.
[6] Edwards, P. (1967). Models and Analogy in Science. In (Ed.), *the Encyclopedia of philosophy.* (pp. 354-359). New York: The Macmillan Company and the Free Press.
[7] Giddens, A. (1984). *The Constitution of Society*. Cambridge: Polity Press.
[8] Goldkuhl, G. & Lyytinen, K. (1982). A Language Acton View of Information Systems. In M. Ginzberg & C. Ross (Eds.). *Proceedings of the Third International Conference on Information Systems* (pp. 13-30). MI: Ann Arbor.
[9] Kahn, B. S. (1982). A Structured Logical Database Design Methodology. In S.Yao, et al. (Eds.), *Database Design Techniques I: Requirements and Logical Structures* (pp. 31-55).Berlin: Springer-Verlag.
[10] Kay. A. (1984). Software. *Scientific American, 251* (3), 41-47.
[11] Kent, W. (1983). Fact-based Analysis and Design. In C. Davis, et al. (Eds.), *Proceedings of the Entity-Relationship Approach to Software Engineering* (pp. 3-53). Amsterdam: Elsevier.
[12] Klein, H. K., & Hirschheim, R. A. (1987). A Comparative Framework of Data Modelling Paradigms and Approaches. *The Computer Journal, 30* (1), 8-15.
[13] Mitroff, I., & Linstone, H. (1993). *The unbounded mind, breaking the chains of traditional business thinking.* New York: Oxford University Press.
[14] Olle, T. William. (1993). Data Modelling and Conceptual Modelling: A comparative analysis of functionality and roles. *Australian Journal of Information Systems*. 1 (1), 46-57.
[15] Polanyi, M. (1973). *Personal knowledge*. London, UK: Routledge & Kegan Paul.
[16] Ryles, G. (1949). *The Concept of Mind*. Harmondsworth: Penguin Books.
[17] von Wittgenstein, L. (1992). Philosophical Investigations. Oxford: Basil Blackwell.